Université de Neuchâtel
Institut de Mathématiques
Institut d'Informatique

SEMINAIRE MATHEMATIQUES ET SOCIETE

Mercredi 9 décembre 2009

(Jean-Léon Gérôme Une collaboration Corneille et Molière – 1874)

Qui a écrit Dom Juan ?

Molière est-il l'auteur des pièces parues sous son nom ?

Dominique Labbé

(Institut d'Etudes Politiques de Grenoble)

dominique.labbe@iep-grenoble.fr

Résumé :

A l'occasion de la parution de son ouvrage "Si deux et deux sont quatre, Molière n'a pas écrit Dom Juan" (Editions Max Milo), Dominique Labbé expose les raisons pour lesquelles Molière n'a pas composé les pièces présentées sous son nom. Après avoir discuté les méthodes statistiques qui attribuent à Corneille les comédies les plus connues de Molière, il explique pourquoi les deux hommes ont collaboré et il montre que cette collaboration n'avait rien d'exceptionnelle.

Nos travaux sur Corneille et Molière appartiennent à un courant qui existe surtout dans les pays anglo-saxons et qui consiste à utiliser des méthodes statistiques pour attribuer à un auteur connu des textes douteux ou d'origine inconnue. Cette "attribution d'auteur" est un aspect particulier d'un problème plus général connu des statisticiens sous le nom de "classification" : la recherche des meilleurs groupements possibles dans les vastes populations, ici des textes. Nous proposons une solution originale consistant en un calcul de distance entre textes, combiné avec plusieurs classifications automatiques.

Cette méthode – qui va être présentée plus en détail - indique que Corneille a probablement écrit toutes les grandes pièces en vers présentées sous le nom de Molière mais aussi *Dom Juan*, l'*Avare* et les 4/5e du *Bourgeois gentilhomme* et du *Malade imaginaire*. A ce propos, il est important de rappeler qu'avant 2001, au moins trois personnes - P. Louÿs, H. Poulaille et H. Wouters - avaient relevé de nombreuses ressemblances troublantes entre les œuvres de Corneille et celles présentées par Molière. Ces trois personnes doivent être considérées comme les "découvreurs", notre travail n'étant qu'une modeste contribution à la solution de cette énigme.

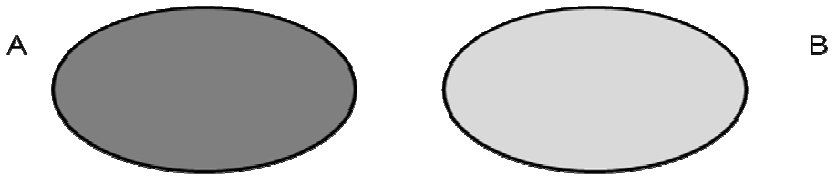
Si notre travail est d'abord d'ordre statistique, il se place aussi sur le terrain historique auquel est consacré l'essentiel de notre dernier livre et qu'on abordera à la fin de cette conférence.

I. STATISTIQUE ET ATTRIBUTION D'AUTEUR

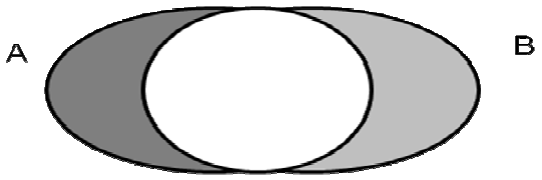
En décembre 2001, C. Labbé et moi-même avons proposé une méthode reposant sur la combinaison d'une mesure de distance avec des classifications que l'on qualifie parfois d'"automatiques", car elles ne nécessitent pas d'intervention humaine.

La distance intertextuelle

La distance entre deux textes se mesure comme la distance séparant deux objets dans l'espace. L'unité de mesure est ici le "mot". Soit deux textes A et B



La distance intertexuelle s'obtient en superposant ces deux textes et en comptant le nombre de mots différents (zones grisées dans le schéma ci-dessous).



Le calcul est opéré de la manière suivante. Soit :

— N_a et N_b longueurs de A et B en mots ("tokens") ;

— V_a et V_b le nombre de "mots différents" ou "vocables" ("types" : c'est l'étendue du vocabulaire) ;

— F_{ia} et F_{ib} le nombre d'apparitions d'un vocable i dans les textes A et B.

Si les deux textes ont la même longueur, leur distance absolue est :

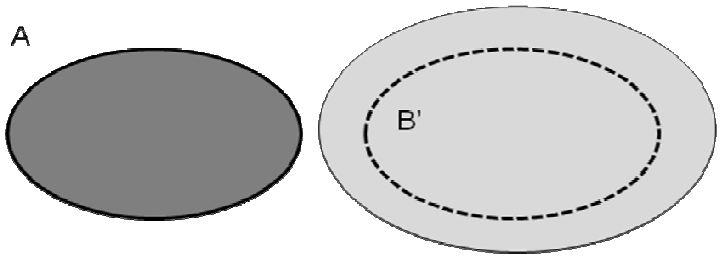
$$(1)\ D_{(A,B)} = \sum_{i \in (A,B)} \left| F_{ia} - F_{ib} \right| \textit{ avec } N_a = N_b$$

Et leur distance relative :

$$(2)\ \mathrm{Drel}_{(A,\,B)} = \frac{\sum_{i\in(A,B)} \left| F_{ia} - F_{ib} \right|}{N_a + N_b}$$

Pour tous les vocables appartenant à A et/ou B, on calcule la différence absolue de leurs fréquences et on fait la somme de ces différences. Cette somme possède les propriétés d'une "distance euclidienne" (longueur du segment de droite unissant deux points). Cet adjectif signifie "conforme à la géométrie d'Euclide" (par un point il ne passe qu'une parallèle à une droite située hors de ce point). Les propriétés d'une distance euclidienne sont : l'identité (la distance d'un point à lui-même est nulle), la symétrie (le résultat est le même que l'on mesure AB ou BA) et l'inégalité triangulaire (le chemin direct entre A et B est toujours plus court qu'en passant par un point C non situé sur le segment AB). Ces propriétés permettent d'opérer des classifications fiables.

Comment procéder quand les deux textes ont des longueurs différentes, tout en conservant ces propriétés ? Supposons que B soit plus long que A ($N_a < N_b$). Nous proposons de "réduire" B à la longueur de A, puis de superposer cette "réduction" (B') sur le texte A.



Considérons un vocable i de fréquence F_{ib} dans B. Son "espérance mathématique" dans B' est :

$$E_{ia(u)} = F_{ib} * U \quad avec \quad U = \frac{N_a}{N_b}$$

Dans la formule (1), en remplaçant la fréquence de chacun des vocables de B par cette espérance mathématique, on obtient la distance intertextuelle absolue :

$$(3)\ \mathrm{D}_{(A,\,B)} = \sum_{i\in(A,B)} \left| F_{ia} - E_{ia(u)} \right|$$

La distance relative est :

$$(4)\ \mathit{Drel}_{(A,\,B)} = \frac{\sum_{i\in(A,B)} \left| F_{ia} - E_{ia(u)} \right|}{\sum_{i\in A} F_{ia} + \sum_{i\in B} E_{ia(u)}}$$

Cette dernière mesure varie uniformément entre un minimum de zéro — distance nulle : mêmes vocables et mêmes fréquences relatives — et un maximum de 1 : tous les mots sont différents. Elle peut aussi être exprimée en %, ‰, etc. Cette mesure relative permet de comparer les résultats obtenus sur un grand nombre de textes de longueurs différentes.

Remarques et limites du calcul

A ce stade de l'exposé, trois remarques s'imposent.

— ce calcul exige une stricte standardisation des graphies. Sinon les textes seront classés selon les conventions graphiques différentes auxquelles ils obéissent (par exemple, les majuscules

initiales de vers différencieront tous les poèmes de tous les textes en prose). Il faut également lemmatiser les textes car le français est une langue flexionnelle, avec de très nombreuses homographies. En effet, dans tout texte en français, plus du tiers des mots peuvent être rattachés à plus d'une entrée de dictionnaire et ce sont les plus fréquents : le (pronom ou article), être ou avoir (substantifs ou verbes)… De plus, la lemmatisation représente le seul moyen de diminuer les effets parasites des basses fréquences (voir sur ce point : Labbé & Labbé 2003) ;

— en théorie, les différences de longueur sont neutralisées puisque, aux arrondis près, la somme des E_{ia} sera égale à N_a. En fait, l'indice de la distance présente une légère tendance à décroître en fonction de l'allongement de la longueur des textes. Lorsque les différences sont trop importantes, les plus longs textes se rapprochent entre eux et s'éloignent des plus petits. Ceci s'explique par la "spécialisation lexicale" — qui s'élève généralement avec la longueur des textes — et par l'amenuisement de l'apport en mots nouveaux au fur et à mesure que le texte s'allonge. Dans une collection de textes de tailles très différentes, ce sont pratiquement toujours les textes les plus longs qui se trouvent au centre de la distribution et les plus courts qui sont les plus décalés. C'est pourquoi, nous avons recommandé dans notre article de 2001 de ne pas appliquer le calcul sur des textes trop courts (en tous cas, pas moins de 1.000 mots) et de s'en tenir à une échelle inférieure à 1:10 dans les différences de longueur. Ces considérations ne sont pas développées ici car elles ne concernent pas l'expérience sur Corneille et Molière : le plus petit texte attribué à Corneille (*Mélicerte*) dépasse 5 000 mots (à ce niveau, les perturbations dues aux basses fréquences dans les textes courts sont négligeables) et la plus longue pièce (*l'Avare*) en compte 21 000. Ce sont les dimensions idéales.

- la mesure enregistre l'influence de 4 facteurs : le genre, l'auteur, l'époque et le thème (pour l'influence respective de chacun de ces facteurs, voir par exemple : Labbé, 2007a et Labbé 2009). Dès lors, pour déterminer l'auteur d'un texte d'origine douteuse ou inconnue, il suffit de le confronter à d'autres – dont l'origine n'est pas douteuse - écrits dans un même genre et à la même époque.

La méthode

La méthode d'attribution d'auteur a été mise au point grâce à un très grand nombre d'expériences réalisées dans des conditions draconiennes. En voici un exemple.

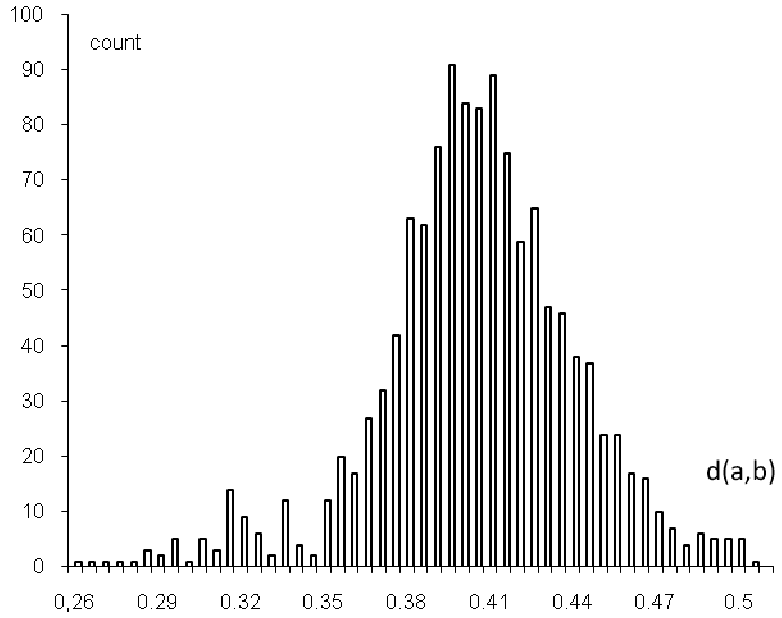
A la demande de Gerard Ledger et de Thomas Merriam, une série d'expériences "en aveugle" ont été réalisées sous leur contrôle. Ils ont relu le compte rendu et en ont accepté la publication (Labbé 2007a).

Lors de la première expérience, G. Ledger a soumis 52 textes anglais anonymés – le "corpus Oxquarry" - en demandant lesquels étaient écrits par les mêmes auteurs. Il avait choisi certains textes parce qu'il semblait difficile de les distinguer bien qu'ayant été écrits par des auteurs différents.

G. Ledger et T. Merriam ont fourni plusieurs indications : les textes sont d'un même genre et à peu près contemporains – il s'agit de romans de la seconde moitié du XIXe — il y a plusieurs auteurs et chacun de ces auteurs a au moins deux textes. Ces textes ont été traités selon la norme "OCP" (Hockey & Martin 1988). Leurs distances mutuelles calculées avec la formule (2) sont rangées par valeurs croissantes dans des classes d'intervalles égaux (figure 1).

Cet histogramme rappelle la courbe en cloche ("Laplace-Gauss"). Mais il y a des anomalies aux deux extrémités, comme si cette figure résultait du mélange de trois populations inégales, chacune distribuées selon une courbe en cloche. Deux sont peu nombreuses et encadrent la principale. La première est composée de distances "anormalement" petites (à gauche), la troisième est constituée des distances les plus grandes (à droite).

Figure 1. Histogramme des distances rangées par valeurs croissantes et intervalles de classe égaux (corpus Oxquarry)



On formule deux hypothèses :

- si les textes de mêmes auteurs sont séparés par les distances les plus courtes, elles constituent la population à gauche du diagramme. Par convention, on fixe la limite à la moyenne diminuée de deux écarts types : soit 70 paires de textes "trop" proches par rapport à la moyenne du corpus.

- si les textes d'auteurs différents sont séparés par les distances les plus longues, elles se trouvent à droite du diagramme (limite : moyenne augmentée de deux écarts types, soit 26 distances "trop" élevées par rapport à la moyenne).

Ces 96 couples (soit 7% du total des couples) permettent d'attribuer 47 textes sur 52 (tableau 1). Après l'expérience, T. Merriam et G. Ledger nous ont livré les noms des auteurs (dernière colonne à droite) : il n'y a aucune erreur. Les deux hypothèses sont donc vérifiées (les distances les plus faibles désignent toujours un même auteur et les distances les plus fortes se rencontrent toujours entre auteurs différents). De nombreuses expériences comme celle-ci ont été réalisées. Jusqu'à ce jour, elles ont toutes validé ces deux hypothèses.

Tableau 1. Détermination des auteurs grâce aux distances remarquables

N°	Textes	Auteurs
1	1D , 2W, 1K, 2R, 1N, 2E, 1F	Stevenson
2	1S, 2M, 2I, 1C	Morris
3	1E, 1Q, 1B, 2A	Butler
4	2L, 2T, 2J, 2X, 1L, 2D, 1X, 2F	Hardy
5	1H, 2Z, 2S	Hardy
6	2O, 2P, 2V	Forster
7	1Z, 2C, 1Y	Tressel
8	1P, 2U, 1R	Chesterton
9	2H, 1G, 1O	Conrad
10	2N, 2Q, 2K	Conrad
11	1M, 2Y, 1W, 2G	Orczy
12	1J, 2B	Morris

Appliquée au corpus Corneille-Molière, cette même méthode classe correctement la plupart des textes, mais elle apporte deux surprises :

- deux comédies en alexandrins de P. Corneille - *le Menteur* (1642) et *la suite du Menteur* (1643) – sont classées avec 14 comédies présentées par Molière entre 1659 et 1673 : douze comédies en alexandrins et deux comédies en prose (*Dom Juan* et l'*Avare*),

- deux comédies "sérieuses" présentées par Molière, également en alexandrins, (*Dom Garcie* et *Psyché*) sont classées avec les tragédies et les comédies héroïques contemporaines de P. Corneille.

A ce stade, on peut formuler trois remarques :

Premièrement, la distance intertextuelle a été testée sur des dizaines de milliers de textes, sans jamais rencontrer de tels croisements entre deux œuvres d'auteurs différents, sauf… quand les deux auteurs supposés ne font qu'un, comme dans le cas de Gary et Ajar (Labbé 2004a et Lafon et Peters 2006), ou quand la même plume de l'ombre a travaillé pour deux employeurs différents (Monière et Labbé, 2006).

Deuxièmement, en répétant des expériences, comme celle organisée avec T. Merriam et G. Ledger, sur un très grand nombre de textes variés, nous avons pu étalonner empiriquement une échelle des distances pour le français. Si l'on exprime les valeurs en "pour 10 000 mots"[1] :

- au-dessous de 2 000 (un mot sur 5) : les deux textes sont écrits dans un même genre, par un seul auteur, à la même époque et sur des thèmes proches,
- entre 2 000 et 2 500 : l'auteur et le genre sont identiques mais les dates de composition et les thèmes sont plus éloignés,
- entre 2 500 et 3 500, soit l'auteur est le même et alors plusieurs facteurs ont changé, soit ce sont deux auteurs différents, travaillant à la même époque dans un même genre sur des thèmes très proches,
- au-dessus de 3 500, si les deux textes sont dans un même genre, alors les auteurs sont différents. Si l'auteur est le même, alors les genres sont certainement différents.

Appliquée au corpus Corneille-Molière, cette échelle confirme la proximité remarquable entre les deux menteurs de P. Corneille et les principales pièces présentées par Corneille (annexe 3).

Le XVIIe siècle fournit des contre-épreuves intéressantes. Par exemple, Corneille et Racine ont écrit chacun une tragédie sur l'amour impossible entre un empereur romain (Titus) et une reine orientale (Bérénice). Ils l'ont fait au même moment, en "aveugle", en alexandrins et en respectant les fameuses "règles" qui enserraient la création théâtrale, spécialement la tragédie. Et tous deux avaient en tête la même référence : la liaison entre Louis XIV et Henriette d'Angleterre, sa belle sœur. Le lieu de l'action et les personnages étaient donc identiques ainsi que les obstacles à cet amour partagé (d'où un vocabulaire commun important…) L'annexe 5 présente le résultat du calcul. La distance entre ces deux pièces (0,256) est plus élevée que toutes celles constatées entre les pièces en vers de Molière et les deux *Menteurs* de Corneille ou entre *Dom Garcie* et les comédies héroïques de Corneille, alors que toutes ces pièces sont séparées par un laps de temps important, que les thèmes sont toujours différents et que le genre "comédie" en vers est nettement moins contraignant que celui de la grande tragédie en alexandrins.

Le XVIIe fournit d'autres exemples. La dernière colonne de l'annexe 3 montre que toutes les distances entre les *Plaideurs* (unique comédie de Racine) et toutes les comédies présentées par Molière sont supérieures à .25 alors qu'elles sont contemporaines. Les *Sophonisbe* de Mairet et de Corneille se distinguent très bien alors qu'ils ont travaillé dans le même genre, sur le même thème, les mêmes événements, les mêmes personnages, en suivant la même trame narrative. Quand Quinault, Corneille et Molière travaillent ensemble sur *Psyché*, Quinault se distingue très

[1] Afin de souligner que cette échelle est valable pour les textes dont les longueurs sont en moyenne de 10 000 mots et comprises entre 4 000 et 25 000 mots environ, comme cela est indiqué dans notre article de 2001.

bien des deux autres… qui sont "trop" proches pour être deux auteurs différents. Les comédies de Quinault se distinguent très bien de celles de Corneille et de Molière, etc.

Conclusion de toutes ces expériences : les distances entre des textes appartenant à un même genre et écrits *par un même auteur* à une même époque, *sur des thèmes différents*, sont systématiquement et nettement inférieures à celles séparant deux auteurs différents, *même quand ils écrivent au même moment, sur un thème identique.*

Enfin, troisième remarque : les grands tableaux de chiffres sont lourds à manipuler manuellement et l'on peut passer à côté d'informations importantes. Des automates peuvent faire ce travail plus efficacement et offrir des représentations graphiques plus simples à consulter.

II. QUELQUES CONFIRMATIONS

Les classifications confirment les anomalies signalées ci-dessus à propos de Corneille et Molière. Deux autres études statistiques vont aider à conclure.

Classifications

L'objectif est double :

- rechercher les meilleurs groupements possibles sans intervention humaine. Deux critères sont utilisés. D'une part, les distances entre les individus composant un même groupe doivent être les plus courtes possibles ; d'autre part, les distances séparant les différents groupes ainsi constitués, doivent être les plus grandes possibles (pour une présentation de la question : Sneath & Sokal 1973 et Benzecri 1980).

- offrir la meilleure représentation possible, en deux dimensions, de ce meilleur ordre possible alors que celui-ci comporte un grand nombre de dimensions (1326 dans l'expérience Oxquarry).

Il existe un grand nombre de méthodes de classification : classification automatique, analyse en composantes principales, analyse factorielle… La méthode dite de la "classification arborée" est classique en génétique (Felsenstein 2004) ou en linguistique historique (Embleton 1986, Holm 2007). Elle repose sur la propriété suivante : si toutes les distances séparant les individus étudiés sont euclidiennes, il existe un "arbre" qui représente exactement les positions respectives de ces individus les uns par rapport aux autres et les meilleurs groupements possibles[2].

Appliquée à l'énigme posée par G. Ledger et T. Merriam (corpus Oxquarry), cette méthode donne la figure 2.

Dans cette figure, les **feuilles** terminales figurent chacune un texte ; les **nœuds** intermédiaires figurent les meilleurs groupements possibles, c'est-à-dire ceux pour lesquels les distances entre les éléments qui composent le groupe sont les plus faibles possible et les distances les séparant des autres les plus grandes possibles. Les segments de droite, ou **arêtes,** sont des **branches** quand elles relient des feuilles à des nœuds et des **troncs** quand elles relient des nœuds entre eux. La distance entre deux points quelconques est figurée par le **chemin** unissant ces points et la longueur de ce chemin est proportionnelle à la distance originelle correspondante (sur la construction de ces arbres, voir Labbé et Labbé 2006 et 2008).

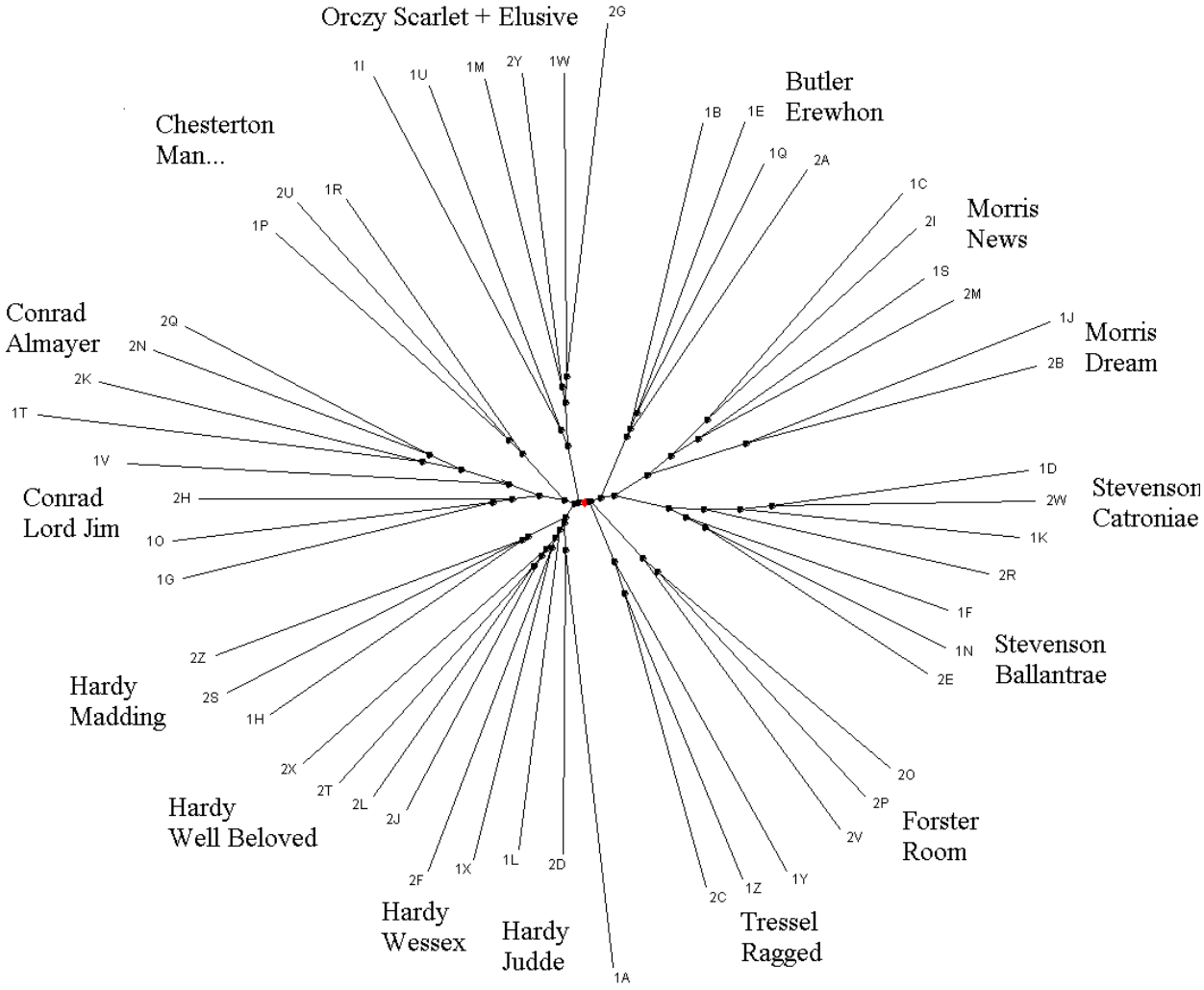
Dans le graphique ci-dessus, tous les textes sont correctement attribués à leurs auteurs et aux œuvres respectives. Cet outil est évidemment plus puissant que l'examen manuel des distances. Il soulève deux questions. Quelle est la précision de la méthode ? Quel risque d'erreur ?

Pour la première question : dans l'arbre ci-dessous, les longueurs des branches sont égales aux distances correspondantes avec une incertitude moyenne de ± 2% (voir Labbé et Labbé 2008). La représentation graphique est donc fiable.

---

[2] Algorithme X. Luong (1988 et 1994). Le programme utilisé pour tracer la figure 2 est présenté dans Rulhman 2003.

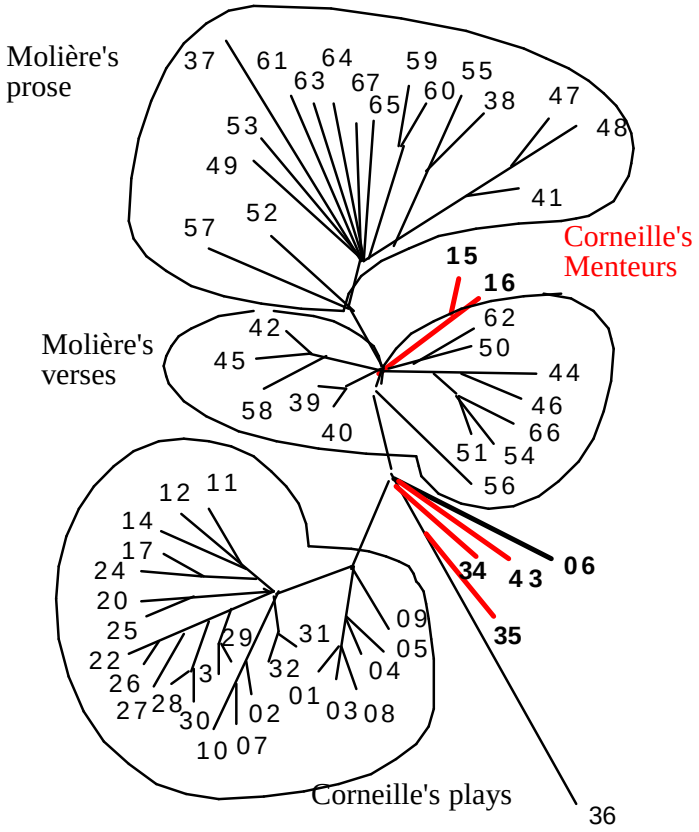
Figure 2. Classification arborée du corpus Oxquarry



NB : les noms des auteurs et des œuvres ont été ajoutés après l'expérience.

Pour la seconde question, il n'y a pas de « risque d'erreur » à proprement parler, puisqu'on ne travaille pas sur échantillons mais sur des populations entières. En revanche, on ne peut écarter, *a priori*, le risque d'un "faux positif". Quelle est la probabilité d'un tel événement ? Cela revient à savoir combien de chances a-t-on de réussir ce genre d'épreuves par hasard. Au numérateur, il y a une seule bonne combinaison ; au dénominateur, on place le nombre de combinaisons différentes que l'on peut réaliser avec ces 52 textes. En théorie, ce nombre est égal à 52 ! (factorielle de 52), soit : 8,07^{e}67 combinaisons différentes. Toutefois, le nombre réel est plus petit. D'une part, tout auteur a au moins deux textes : il faut donc éliminer les combinaisons comportant des "célibataires". D'autre part, la distance intertextuelle étant symétrique, il faut diviser par deux le nombre de couples possibles, par 6 le nombre de trios, par 8 le nombre de quatuors, etc. Cela donne tout de même 6,1^{e}19 combinaisons possibles. Nous avions donc une chance sur 61 millions de billions de réussir cette épreuve par hasard (le faux positif). Ceux qui prétendent que ce genre d'exercice est "facile" devraient y réfléchir. Cette remarque vaut aussi pour la classification effectuée sur Corneille et Molière par X. Luong en 1999 (figure 3).

Figure 3. Classification arborée sur l'ensemble des œuvres théâtrales de Corneille et Molière



Ce graphe a été tracé par M. X. Luong de l'université de Nice. Pour les numéros des pièces, se reporter à l'annexe 1. Les traits en gras correspondent à : n° 06 Corneille : *Comédie des Tuileries* (écrite par Corneille pour Richelieu en 1634) ; n° 15 et 16 Corneille : *Le Menteur* et la *Suite du Menteur* (1642 et 1643) ; n° 34 : passages de *Psyché* écrits par Corneille ; n° 35 : passages de *Psyché* attribués par l'éditeur à Molière ; n° 36 : prologue de *Psyché* écrit par Quinault ; n° 43 : *Dom Garcie* présenté par Molière

Cet arbre isole clairement trois ensembles : en bas, les pièces de Corneille, au milieu les pièces en vers présentées par Molière et en haut, avec une dispersion plus grande, les pièces en prose. On retrouve les "anomalies" suivantes :

— en bas, les contributions respectives de Corneille et de Molière à une même pièce, *Psyché*, se rejoignent avec *Dom Garcie de Navarre* (présenté par Molière) et avec le troisième acte de la *Comédie des Tuileries*, écrite par Corneille pour le Cardinal Richelieu en… 1634 (soit 37 ans avant *Psyché*) !

— au milieu des pièces en vers de Molière, figurent les deux *Menteurs* de Corneille. La chose est d'autant plus surprenante que les *Menteurs* ont été créés 15 ans avant *l'Etourdi* (première pièce en vers de Molière) et 30 ans avant les *Femmes savantes*.

Ceci confirme donc les déductions tirées de l'examen direct de la matrice des distances : *Psyché* et *Dom Garcie* (deux comédies héroïques présentées par Molière) sont anormalement proches de celles de P. Corneille ; toutes les autres comédies en vers présentées par Molière auxquelles viennent s'ajouter *Dom Juan* et l'*Avare* sont aussi anormalement proches des deux *Menteurs*, de Corneille.

Dans l'histoire littéraire française de ces quatre derniers siècles, on ne trouve aucun autre exemple de double croisement entre deux œuvres, du moins quand les deux auteurs sont effectivement différents. Cela est confirmé par deux autres indices.

Deux autres indices

Les combinaisons de mots les plus fréquents employées par Corneille et Molière révèlent les mêmes proximités entre ces deux auteurs supposés (tableau 2).

Tableau 2. Les combinaisons "pseudo-auxiliaire + infinitif" chez Corneille, Molière et Racine (fréquence pour 100.000 mots)

P. Corneille		Molière		Racine	
Syntagmes	F	Syntagmes	F	Syntagmes	F
faire voir	***33,8***	***faire voir***	***31,5***	aller voir	12,0
pouvoir être	***18,8***	***pouvoir être***	***25,5***	pouvoir voir	**9,6**
pouvoir faire	***18,4***	***pouvoir faire***	***25,5***	faire entendre	9,0
faire naître	13,9	vouloir dire	24,9	pouvoir faire	8,4
pouvoir voir	**13,4**	***vouloir faire***	***19,5***	aller chercher	7,8
devoir être	12,7	pouvoir dire	14,5	faire parler	7,8
pouvoir souffrir	10,8	pouvoir avoir	13,7	pouvoir être	7,8
vouloir faire	***9,9***	aller faire	13,2	venir chercher	7,2
faire connaître	9,6	avoir faire	13,2	faire éclater	6,6
devoir faire	8,7	**pouvoir voir**	**12,3**	falloir partir	6,6

Racine ne partage avec Corneille et Molière que trois combinaisons (soulignées) : "pouvoir voir", "pouvoir faire" et "pouvoir être", mais avec un classement et des densités très différentes. En revanche, Corneille et Molière en ont cinq en commun (en gras) dont les trois premières dans le même ordre et avec des densités voisines (en italiques). Etant donné le nombre des combinaisons possibles, la probabilité pour qu'une telle "coïncidence" survienne au hasard est infinitésimale.

A notre connaissance, il existe un seul cas comparable dans les 4 derniers siècles de littérature française (tableau 3). Depuis huit ans, personne n'a pu en trouver un autre concernant deux auteurs différents. Il faut donc admettre que le cas Molière-Corneille est unique dans l'histoire littéraire…

Tableau 3. Les principaux groupes verbaux chez Ajar et Gary (fréquence pour 10.000 mots)

Ajar	F	Gary	F
vouloir dire	7,20	**vouloir dire**	5,80
pouvoir être	4,90	**pouvoir être**	3,62
pouvoir faire	4,40	**pouvoir faire**	2,97
devoir être	3,10	**devoir être**	2,39
aller faire	2,90	**aller faire**	2,20
pouvoir vivre	2,70	**devoir avoir**	2,13
laisser tomber	2,30	pouvoir dire	2,13
devoir avoir	2,10	**laisser tomber**	1,87
aller voir	2,00	laisser aller	1,55
devoir faire	1,80	faire passer	1,49

Autre indice intéressant : le sens spécifique que chaque auteur donne aux principaux mots qu'il emploie. Grâce à l'étude des réseaux sémantiques, nous pouvons affirmer que, chez Corneille et Molière, les principaux vocables ont le même sens, ou plutôt, que ceux de Molière s'inscrivent comme un sous-ensemble dans ceux de Corneille. Le plus évocateur est le mot "amour" – substantif le plus employé par Corneille comme par Molière. Par exemple, "haine" est le vocable le plus fortement associé à l'amour dans les deux œuvres : *la haine détruit l'amour*, *l'amour se transforme en haine*, *l'amour résiste à la haine*… Les vocables les plus étrangers à l'amour sont : *Dieu, père*, *frère*, *fils*, *gendre*… (pour une présentation détaillée des calculs et de l'univers de l'amour chez Corneille : Labbé & Labbé 2006). Là encore, ces significations sont propres à Corneille et ne se retrouvent pas chez ses contemporains. Nous avons refait cette démonstration à Louvain-la-

Neuve devant deux de nos contradicteurs et une centaine de spécialistes de la statistique appliquée au langage (Labbé 2004b). La démonstration est en ligne et n'a pas été contredite.

III. UN PEU D'HISTOIRE

Naturellement, une expérience d'attribution d'auteur n'a de sens que si, historiquement, la collaboration était possible et qu'il n'y a pas d'obstacle à celle-ci. Dans l'article de 2001, nous avons indiqué que c'est bien le cas pour Corneille et Molière et nous avons signalé que, dès le début des succès de Molière à Paris, des rumeurs ont circulé sur la paternité des œuvres qu'il présentait. Mon essai – *Si deux et deux sont quatre* - examine en détail ce volet historique, spécialement en ce qui concerne la scène théâtrale du temps. Cette lecture peut être utilement complétée par celle de Denis Boissier (2008) dont les conclusions recoupent très largement les nôtres.

Le théâtre au XVIIe

La clef essentielle est fournie par l'examen de la production théâtrale du temps[3]. En effet, durant la seconde moitié du XVIIe siècle, 6 pièces sur 10 ont été présentées par des comédiens et non pas par les écrivains qui les ont composées. C'est le cas de 9 comédies sur 10. Notre essai évoque les principaux "comédiens poètes", confrères de Molière : Baron, Brécourt, Champmeslé, Dancourt, Dorimond, Hauteroche, La Thuilerie, Montfleury, Poisson, Raisin, Rosimond, Villiers… En effet, il s'agit d'un véritable système, qui donne son titre à une comédie présentée, en 1673, par la troupe de Molière sous le nom de Montfleury alors qu'elle est de… Thomas Corneille, le petit frère de Pierre. En effet, T. Corneille, comme les autres grands écrivains du temps, présentait des pièces sérieuses sous son propre nom et des comédies sous le nom de Montfleury puis de Hauteroche. Notre essai donne d'autres exemples de collaborations semblables : Boursault avec le comédien Poisson ou La Fontaine avec Champmeslé.

Notre essai expose les raisons d'un tel système. Les principales sont les suivantes :

- les comédies satiriques, comme celles de Molière, plaisaient au public parisien – qui était assez frondeur – elles étaient nécessaires à l'équilibre économique des troupes mais elles étaient condamnées par l'Eglise, par une partie de la Cour et par l'Académie. Les écrivains célèbres qui composaient ces comédies préféraient donc rester dans l'ombre,

- il n'y avait pas de propriété intellectuelle et les troupes n'avaient pas la personnalité juridique. Pour garder le contrôle sur ces textes, il fallait donc qu'un comédien achète la pièce à l'écrivain et assure son exploitation,

- le comédien poète présentait la pièce à l'assemblée de la troupe, il avançait une partie des sommes nécessaires pour les décors, la musique, les ballets, il surveillait la distribution des rôles, les costumes, la mise en scène, il veillait à ce que la pièce ne soit pas retirée trop vite de l'affiche, il surveillait le partage de la recette qui avait lieu chaque soir après la représentation, il négociait la publication avec les éditeurs et… il encaissait les moqueries et les critiques à la place de l'écrivain.

Et Molière ?

Molière était-il différent de ses confrères comédiens et auteurs pour la galerie ? Etait-il un grand écrivain comme on le croit aujourd'hui ?

Les documents de l'époque montrent que Molière était un riche financier qui se fait passer pour noble et serviteur du roi. Dans ses deux maisons, richement meublées, il n'y avait pas de bibliothèque ni de bureau, très peu de livres. Molière ne laisse aucun manuscrit, aucune correspondance. Personne ne l'a vu écrire. D'ailleurs, grâce au registre tenu par La Grange, un

---

[3] Ce recensement est rendu possible grâce à un site en ligne anglais ("Cesar") qui présente toutes les pièces de théâtre connues des XVIIe et XVIIIe siècle français avec leur date de création, leur "auteur" et les comptes rendus.

des comédiens de la troupe, on connaît son emploi du temps. Celui-ci ne lui donnait pas le loisir nécessaire pour composer l'œuvre qu'on lui prête…

Les deux principaux critiques de l'époque ont clairement indiqué qu'il n'était pas l'auteur des pièces qu'il présentait. Par exemple, en 1663, Robinet du Laurens a écrit : "On ne peut pas dire que Molière soit une source vive mais seulement un bassin qui reçoit ses eaux d'ailleurs pour ne point le traiter plus mal en le comprenant dans la comparaison que quelques-uns ont faite (…) avec des ânes seulement capables de porter des grands fardeaux". Donneau de Visé, également en 1663 : "Ses enfants ont plus d'un père (…) Tout le Parnasse s'assemble quand il veut faire quelque chose".

Les écrivains contemporains n'ont guère été plus tendres. Après avoir fait la connaissance de Molière, en 1663, La Fontaine n'a plus écrit une ligne en sa faveur. Boileau a composé plusieurs satires dans lesquelles il met en doute la qualité d'auteur de Molière, avant d'écrire en 1674, qu'il est "un faux plaisant, à grossière équivoque qui, pour me divertir n'a que la saleté". Mme de Sévigné ouvrait sa porte à de nombreux artistes, mais elle n'a jamais reçu Molière chez elle, alors qu'elle citait souvent ses pièces…

Plusieurs contemporains – dont Boileau et l'Abbé d'Aubignac - ont clairement désigné les frères Corneille comme étant des "plumes mercenaires".

A trois reprises P. Corneille a été explicitement désigné comme étant l'auteur de trois pièces présentées par Molière. En 1662, l'éditeur Quinet – qui était l'un des éditeurs de Corneille – indique que le *Dépit amoureux* est de sa main. En 1670, Robinet écrit dans *Gazette rimée* du 22 novembre : "On donne de Corneille sa *Bérénice* sans pareille sur le théâtre de Molière (…) et son charmant *Bourgeois gentilhomme*". En 1671, l'éditeur de *Psyché* indique que cette pièce a été versifiée au trois quarts par P. Corneille.

Enfin, on peut signaler quelques faits étonnants. En 1658, à 35 ans, Molière n'avait encore à son actif que trois méchantes farces. Alors, il séjourne quelques mois à Rouen, ville où résidaient les frères Corneille qui l'introduisent à la cour de Louis XIV. A partir de ce moment, Molière se met à produire des chefs d'œuvre, spécialement à partir de 1662, date de l'installation à Paris des frères Corneille. Quant à ceux-ci, ils semblent avoir fait face aisément, au moins jusqu'en 1673, aux charges énormes nécessaires pour établir leurs dix enfants alors que leurs revenus connus ne pouvaient y suffire.

CONCLUSIONS

Tous nos articles sur cette question sont consultables sur le site "archives en ligne" du CNRS (HAL-SHS). Nos programmes et nos données sont dans le domaine public. La plupart des documents historiques que nous avons cités sont en ligne, les autres ont été publiés. Tout est donc vérifiable. Enfin, depuis huit ans, nous proposons à nos contradicteurs de réaliser les expériences de leur choix.

Quelles conclusions tirer de ce faisceau d'indices précis, sérieux, concordants – et tous vérifiables ?

Du volet statistique, on conclut qu'il existe entre les œuvres de Corneille et les principales pièces présentées par Molière des proximités uniques dans l'histoire littéraire française (pour deux écrivains contemporains, travaillant dans le même genre, sur des thèmes proches).

Du volet historique, on conclut que cette proximité est logique puisque Corneille a fait comme les autres écrivains de son temps : il a présenté ses tragédies sous son nom et ses comédies sous le nom d'un comédien poète. C'est l'oubli de ce système du comédien poète qui a permis à la légende du Molière grand auteur de prendre corps, bien après sa mort, dans des circonstances évoquées dans mon essai.

Dès lors, il faut admettre que P. Corneille a écrit toutes les pièces en vers présentées sous le nom de Molière ainsi que l'*Avare* et *Dom Juan* et au moins 4 actes sur 5 du *Bourgeois gentilhomme* et du *Malade imaginaire*.

Certains littéraires considèrent que cette conclusion ne change rien puisque le texte se suffit à lui-même et qu'il n'est pas nécessaire de connaître l'auteur pour l'interpréter. Dans ce cas, il faut cesser de chercher dans la vie imaginée de Molière des clefs d'interprétation, comme le font tous les "Moliéristes" (c'est le nom que se donnent les partisans de la thèse du Molière grand écrivain).

Il faut admettre que les pièces présentées sous le nom de Molière répondaient aux attentes des comédiens et du public parisien, qu'elles s'inscrivaient dans l'atmosphère intellectuelle de l'époque... Il faut aussi admettre que l'œuvre des frères Corneille est à redécouvrir à la lumière de leur activité de "plumes de l'ombre".

Au-delà de Corneille et Molière, nous voudrions souligner aussi l'apport de la statistique appliquée aux sciences du langage et à l'analyse littéraire.

Pour ce qui concerne l'analyse littéraire, on pourra enfin connaître le vocabulaire d'un auteur, d'une œuvre, d'une époque, mais aussi le sens spécifique des mots, les styles, les thèmes préférés. On décrira les évolutions dans une œuvre, les ruptures, les continuités. On pourra reconstituer les principaux courants littéraires, éclairer les filiations et les influences mutuelles.

Pour les sciences du langage, on peut enfin étudier, de façon scientifique, la manière dont une langue est parlée et écrite par ses usagers. En effet, pour les principales langues européennes, on dispose déjà de vastes échantillons représentatifs oraux et écrits appartenant à de multiples genres et provenant de toutes les classes sociales. Cela permet une étude scientifique de la grammaire, des structures de phrase, du lexique… Ce sont des outils indispensables pour l'enseignement — surtout à ceux dont ce n'est pas la langue maternelle —, les correcteurs orthographiques, la documentation et l'archivage, les traducteurs automatiques... Sans ces outils, une langue est condamnée à plus ou moins court terme à la disparition dans un siècle dominé par la communication. De tels outils existent pour l'anglais, l'allemand, l'espagnol, le tchèque, l'écossais… mais pas pour le français.

Une anecdote pour finir. En 1998, quand nous avons commencé à apercevoir l'étrange parenté des principales pièces de Molière avec celles de Corneille, nous nous sommes documentés, ce qui nous a permis de découvrir que de nombreuses personnes avaient eu cette intuition avant nous. C'est particulièrement le cas du poète P. Louÿs (au début du XXe siècle), du romancier H. Poulaille (dans les années 1950) et d' H. Wouters (dans les années 1990). Il y en a certainement eu d'autres. D. Boissier – animateur de l'excellent site corneille-moliere.org - nous a signalé le peintre J.-L. Gérôme qui a réalisé, en 1873 ou 1874, un tableau (reproduit au début de cette communication) dont le titre est "Corneille Molière : une collaboration". On y voit Corneille lisant un texte à Molière, aussi richement vêtu que Corneille l'est pauvrement…

Au fond, notre seul mérite est d'avoir eu à notre disposition la puissance des ordinateurs modernes et d'internet qui ont beaucoup facilité nos recherches statistiques et historiques.

Références

Tous nos travaux cités sont accessibles en ligne à partir de notre page personnelle. On consultera également le site corneille-moliere.org.

Benzecri J.-P. (1980). *L'analyse des données. 1. La taxinomie*. Paris : Dunod.

Boissier D. (2008). *Tout savoir sur l'affaire Corneille-Molière*. Paris : Association cornélienne de France.

Embleton S. (1986). *Statistics in Historical Linguistics*. Bochum : Brokmeyer.

Felsenstein J. (2004). *Inferring Phylogenies*. Sunderland : Sinauer Ass.

Hockey S. & Martin J. (1988). *OCP Users' Manual*. Oxford. Oxford University Computing Service.

Holm H. J. (2007). "The New Arboretum of Indo-European "Trees". Can New Algorithms Reveal the Phylogeny and Even Prehistory of Indo-European?". *Journal of Quantitative Linguistics*. 14-2, p. 167-214.

Labbé C. & Labbé D. (2001). "Inter-Textual Distance and Authorship Attribution Corneille and Molière". *Journal of Quantitative Linguistics*. 8-3, p. 213-231.

Labbé C. & Labbé D. (2003). "La distance intertextuelle". *Corpus*. 2, p. 95-118.

Labbé C. & Labbé D. (2005). "How to Measure the Meanings of Words ? Amour in Corneille's Work". *Langage Resources Evaluation*. 39, p. 335-351.

Labbé C. & Labbé D. (2006). "A Tool for Literary Studies. Intertextual Distance and Tree Classification. *Literary and Linguistic Computing*. 21-3, p. 311-326.

Labbé C. & Labbé D. (2007a). "Baudelaire, Rimbaud et Verlaine". VIIIe journées de l'ERLA, *Aspects linguistiques du texte poétique*. Brest 16-17 novembre 2007.

Labbé C. et Labbé D. (2007b). *Corneille a écrit 16 pièces représentées sous le nom de Molière. Réponses à : Viprey Jean-Marie et Ledoux Claude-Nicolas, 'About Labbé's "Inter-textual Distance"*. Grenoble : PACTE-IEP.

Labbé C. et Labbé D. (2009). "Existe-t-il un genre épistolaire. Hugo, Flaubert et Maupassant". *Xe journées de l'Erla*. Brest novembre 2009.

Labbé D. (2004a). *Romain Gary et Emile Ajar*. Grenoble : Cerat-IEP, mai 2004.

Labbé D. (2004b). *Corneille et Molière. Table ronde 7e Journées d'Analyse des Données Textuelles*. Louvain-la-Neuve 11 mars 2004. Grenoble : CERAT-IEP

Labbé D. (2007). "Experiments on Authorship Attribution by Intertextual Distance in English". *Journal of Quantitative Linguistics*, April 2007, 14-1. p. 33-80.

Lafon M. & Peeters B. (2006). *Nous est un autre*. Paris, Flammarion.

Luong X. (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Paris : Université de Paris V.

Luong X. (1994). "L'analyse arborée des données textuelles : mode d'emploi". *Travaux du cercle linguistique de Nice*. 16, p. 25-42.

Monière D. & Labbé D. (2006). "L'influence des plumes de l'ombre sur les discours des politiciens". In Condé C. & Viprey J.-M. *Actes des 8e Journées internationales d'Analyse des données textuelles*. Besançon, 19-21 avril 2006, II, p. 687-696.

Ruhlman M. (2003). *Analyse arborée. Représentation par la méthode des groupements*. Grenoble : Polytech' – CERAT.

Wouters H., Ville de Goyet C. de (1990). *Molière ou l'auteur imaginaire ?* Bruxelles : Eds Complexe.

Annexe I. Les pièces de Pierre Corneille.

		Création	Genre	Longueur en mots
1	Mélite	1630 ?	Comédie	16 690
2	Clitandre	1631	Tragi-comédie	14 402
3	La Veuve	1631	Comédie	17 661
4	La Galerie du Palais	1632	Comédie	16 140
5	La Suivante	1633	Comédie	15 160
6	Comédie des Tuileries	1634	Comédie	3 627
7	Médée	1635	Tragédie	14 269
8	La Place Royale	1634	Comédie	13 801
9	L'illusion comique	1636	Comédie	15 428
10	Le Cid	1636	Tragi-comédie	16 677
11	Cinna	1641	Tragédie	16 126
12	Horace	1640	Tragédie	16 482
13	Polyeucte	1641	Tragédie	16 472
14	Pompée	1642	Tragédie	16 492
15	Le Menteur 1	1642	Comédie	16 653
16	Le Menteur 2	1643	Comédie	17 675
17	Rodogune	1644	Tragédie	16 842
18	Théodore	1645	Tragédie	17 121
19	Héraclius	1647	Tragédie	17 433
20	Andromède	1650	Tragédie	15 514
21	Don Sanche	1650	Comédie héroïque	16 947
22	Nicomède	1651	Tragédie	16 923
23	Pertharite	1651	Tragédie	17 121
24	Oedipe	1659	Tragédie	18 618
25	Toison d'Or	1661	Tragédie	20 343
26	Sertorius	1662	Tragédie	17 675
27	Sophonisbe	1663	Tragédie	16 858
28	Othon	1664	Tragédie	16 971
29	Agésilas	1666	Tragédie	18 227
30	Atilla	1667	Tragédie	16 788
31	Tite et Bérénice	1670	Comédie héroïque	16 697
32	Pulchérie	1672	Tragédie	16 630
33	Suréna	1674	Tragédie	16 545
34	Psyché Corneille	1671	Comédie en vers	10 067
35	Psyché Molière	1671	Comédie en vers	4 816
36	Psyché Quinault	1671	Comédie en vers	1 399

Sources : Charles Marty-Laveaux. *Œuvres complètes de P. Corneille.* Paris : Hachette 1862. Collection Les Grands écrivains de la France.

Annexe II.

L'œuvre théâtrale représentée sous le nom de Molière

		Création	Genre	Longueur mots
37	La jalousie	Avant 1659	Comédie prose	3 501
38	Médecin volant	Avant 1659	Comédie prose	3 876
39	L'étourdi*	1659	Comédie vers	18 671
40	Dépit amoureux*	1659	Comédie vers	16 242
41	Précieuses ridicules	1660	Comédie prose	6 648
42	Sganarelle*	1660	Comédie vers	6 042
43	Dom Garcie*	1661	Comédie héroïque vers	17 049
44	L'école des maris*	1661	Comédie vers	10 536
45	Les fâcheux*	1661	Comédie vers	7 922
46	L'école des femmes*	1662	Comédie vers	16 625
47	Critique de l'école	1663	Comédie prose	8 610
48	L'impromptu	1663	Comédie prose	7 168
49	Mariage forcé	1664	Comédie prose	6 058
50	Princesse d'Elide*	1664	Comédie vers et prose	11 333
51	Le Tartuffe*	1664	Comédie vers	18 271
52	Dom Juan*	1665	Comédie prose	17 452
53	L'amour médecin	1665	Comédie prose	6 147
54	Le Misanthrope*	1666	Comédie vers	17 180
55	Médecin malgré lui	1666	Comédie prose	9 317
56	Mélicerte*	1666	Comédie vers	5 540
**	Comédie pastorale	1667	Comédie vers libres	732
57	Le sicilien	1667	Comédie prose	5 375
58	Amphytrion*	1668	Comédie vers libres	15 117
59	Georges Dandin	1668	Comédie prose	11 009
60	L'avare*	1668	Comédie prose	21 033
61	M. de Pourceaugnac	1669	Comédie prose	11 803
62	Amants magnifiques*	1670	Comédie vers & prose	11 983
63	Bourgeois gentilhom.*	1670	Comédie prose	17 132
64	Fourberies de Scapin	1671	Comédie prose	14 245
65	Escarbagnas	1671	Comédie prose	5 564
66	Femmes savantes*	1672	Comédie vers	16 863
67	Malade imaginaire*	1673	Comédie prose	19 919

* Pièce écrite, en tout ou partie, par P. Corneille

** Pièce retirée dans les expériences à cause de sa petite taille.

Sources : Eugène Despois. *Œuvres complètes de Molière.* Paris : Hachette, 1876. Collection Les Grands écrivains de la France.

Annexe III

Distances séparant les deux *Menteurs* (Corneille) et les *Plaideurs* (Racine)
de toutes les pièces de Molière

N°	Pièces	Genre	Le Menteur (Corneille 1642)	Suite du Menteur (Corneille 1643)	Les plaideurs (Racine : 1668)
15	Le Menteur (1642)	Vers	0,000	0,180	0,296
16	La suite du Menteur (1643)	Vers	0,180	0,000	0,293
34	Psyché Corneille (1671)	Vers	0,288	0,273	0,348
36	Psyché Molière (1671)	Vers	0,329	0,325	**0,354**
37	La jalousie du barbouillé (avant 1660)	Prose	0,341	0,331	0,327
38	Médecin volant (avant 1660)	Prose	0,310	0,293	0,302
39	**L'étourdi (1658)**	Vers	**0,205**	**0,206**	**0,269**
40	**Dépit amoureux (1658)**	Vers	**0,215**	**0,212**	**0,270**
41	Précieuses ridicules (1660)	Prose	0,315	0,314	0,314
42	Sganarelle ou le cocu imagin. (1660)	Vers	0,259	0,253	0,293
43	Dom Garcie de Navarre (1661)	Vers	0,280	0,273	0,359
44	**L'école des maris (1661)**	Vers	**0,223**	**0,217**	0,279
45	**Les fâcheux (1661)**	Vers	**0,248**	**0,248**	0,306
46	**L'école des femmes (1662)**	Vers	**0,226**	**0,217**	**0,261**
47	Critique de l'école des femmes (1663)	Prose	0,323	0,319	0,340
48	L'impromptu de Versailles (1663)	Prose	0,321	0,316	0,323
49	Mariage forcé (1664)	Prose	0,322	0,302	0,320
50	**Princesse d'Elide (1664)**	Vers Prose	0,251	**0,243**	0,314
51	**Le Tartuffe (1664)**	Vers	**0,242**	**0,228**	0,275
52	**Dom Juan (1665)**	Prose	0,259	**0,248**	0,281
53	L'amour médecin (1665)	Prose	0,292	0,289	0,287
54	**Le Misanthrope (1666)**	Vers	0,252	**0,234**	0,283
55	Médecin malgré lui (1666)	Prose	0,298	0,289	0,296
56	**Mélicerte (1666)**	Vers	0,257	**0,250**	0,322
57	Le sicilien ou l'amour peintre (1667)	Prose	0,277	0,260	0,301
58	Amphytrion (1668)	Vers libres	0,253	0,256	0,297
59	Georges Dandin (1668)	Prose	0,292	0,279	0,292
60	**L'Avare (1668)**	Prose	0,256	**0,244**	**0,270**
61	M. de Pourceaugnac (1669)	Prose	0,292	0,283	0,285
62	Amants magnifiques (1670)	Prose	0,282	0,279	0,329
63	Bourgeois gentilhomme (1670)	Prose	0,294	0,280	0,286
64	Fourberies de Scapin (1671)	Prose	0,269	0,263	0,281
65	Comtesse d'Escarbagnas (1671)	Prose	0,311	0,300	0,305
66	**Femmes savantes (1672)**	Vers	0,260	**0,248**	0,283
67	Malade imaginaire (1672)	Prose	0,282	0,270	0,278
	Moyenne oeuvre de Molière		*0,275*	*0,266*	*0,299*
	Moyenne pièces en vers de Molière		*0,241*	*0,234*	*0,290*
	Moyenne oeuvre de Corneille		*0,252*	*0,249*	*0,347*
	Moyenne oeuvre de Racine		*0,314*	*0,311*	*0,376*

Annexe IV.

Distances séparant *Dom Garcie* (Molière) et *Psyché* (Corneille et Molière) des dernières pièces de Corneille.

Ultimes pièces de Corneille	Dom Garcie (Molière,1661)	Psyché (Corneille 1671)
Rodogune (1644)	0,245	0,231
Theodore (1645)	0,234	0,245
Heraclius (1647)	0,248	0,273
Andromède (1650)	0,241	**0,218**
DonSanche (1650)	0,224	0,251
Nicomède (1651)	0,244	0,264
Pertharite (1651)	0,235	0,263
Œdipe (1659)	0,223	0,226
Toison d'or (1661)	**0,221**	**0,220**
Sertorius (1662)	0,230	0,238
Sophonisbe (1663)	0,228	0,236
Othon (1664)	0,235	0,240
Agesilas (1666)	0,234	0,233
Attila (1667)	0,235	0,227
Tite et Bérénice (1670)	**0,227**	0,235
Psyché (1671)	***0,230***	—
Pulcherie (1672)	0,230	0,226
Surena (1674)	**0,216**	0,224
Moyenne Corneille	*0,243*	*0,244*
Moyenne Molière	*0,286*	*0,297*

Annexe V

Principales distances caractéristiques entre Corneille et Racine à l'époque de *Tite et Bérénice*.

	Tite et Bérénice (Corneille, 1670)	Bérénice (Racine, 1670)
CORNEILLE :		
Agésilas (1666)	0.159	0.278
Attila (1667)	0.180	0.289
Tite et Bérénice (1670)	0	**0.256**
Pulchérie (1672)	0.155	0.271
Suréna (1672)	0.156	0.264
RACINE :		
Andromaque (1667)	0.259	0.225
Britannicus (1669)	0.251	0.209
Bérénice (1670)	**0.256**	-
Bazajet (1672)	0.262	0.220
Mithridate (1673)	0.249	0.206